Childhood gender-typed behavior and emotional or peer problems:

A prospective birth-cohort study

Anna-Sophia Warren, Kimberley A. Goldsmith, Katharine A. Rimes

Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

Abbreviated Title: Childhood gender-typed behavior and emotional/peer problems

Word count: 6,500

Accepted for publication in Journal of Child Psychology and Psychiatry

https://onlinelibrary.wiley.com/journal/14697610

Abstract

Background

Retrospective and cross-sectional studies often report associations between childhood gender-nonconformity and greater emotional and peer difficulties. This study used the ALSPAC birth-cohort to investigate relationships between childhood gender-typed behavior and peer and emotional problems throughout childhood and adolescence.

Methods

11,192 participants had at least one measure of parent-rated gender-typed behavior in infancy; 7,049 participants had a measure of child-rated gender-typed behavior at 8.5 years. Separate linear mixed regression models were fitted to assess whether parent and child-rated gender-typed behaviors were associated with emotional and peer problems across childhood and adolescence (6 to 16 years old). The effect of adding covariates (self-esteem, abuse, bullying, feeling accepted by peers, and feeling different) on these relationships was assessed.

Results

For boys, more gender-nonconforming behavior was associated with greater subsequent emotional and peer problems in childhood and adolescence. Adjusting for self-esteem, relational bullying victimization, feeling different or feeling accepted by peers reduced some of these associations. In contrast, for girls, more gender-nonconforming behavior was associated with *fewer* emotional problems in childhood and adolescence. In girls, *self*-reported gender-nonconforming behavior was also associated with fewer parent-rated peer problems but *parent*-rated gender nonconforming behavior was associated with more parent-rated peer problems; this latter association was partly explained by self-esteem, bullying and abuse victimization. These associations were statistically significant but small.

Conclusions

Overall, more female-typical behaviors were generally associated with greater subsequent emotional and peer problems, for both boys and girls. Future studies should investigate factors that reduced these associations, as well potential negative effects of female-typical behaviors or advantages of male-typical ones. As this was a 14-year longitudinal study, the relationships between gender-role behaviors and emotional/peer problems warrant further research despite the small association sizes.

Keywords

ALSPAC; gender typicality; sex differences; longitudinal; prospective; psychological problems; relationship problems.

Introduction

Gender expression refers to the way that children express themselves in a stereotypically masculine or feminine way, through their interests, friendships, and clothing choices. This gender-typed behavior in childhood has been shown to be significantly associated with gender-typed behavior in adolescence (effect sizes 0.37-0.46; Golombok , Rust, Zervoulis, Golding, & Hines, 2012). Those who do not follow the social norms expected for their assigned sex at birth are referred to as 'gender nonconforming'.

Retrospective ratings of childhood gender nonconformity (CGN) have been found to be associated with increased levels of mental illness and suicidality in adolescence and adulthood in sexual minority (e.g. Friedman, Koeske, Silvestre, Korr, & Sites, 2006) and heterosexual (Roberts, Rosario, Slopen, Calzo, & Austin, 2013) populations. Cross-sectional studies in young adulthood (e.g. Rieger & Savin-Williams, 2012) have similarly found current gender nonconformity to be associated with poorer mental health. While many studies report similar outcomes for males and females (e.g. Rieger & Savin-Williams, 2012), some studies have found that significant associations between recalled CGN and emotional or relationships difficulties are smaller for females than males (e.g. Roberts et al., 2013), or only present in males (e.g. Lippa, 2008).

The authors are aware of only two prospective studies investigating CGN and subsequent mental illness (both based on the same birth cohort as the present study). One found CGN was not associated with having an anxiety disorder at 17 years after adjustment for sexual orientation (Jones, Robinson, Oginni, Rahman, & Rimes, 2017). However, direct associations between CGN and anxiety disorders without prior adjustment for sexual orientation were not reported. The other (Oginni, Robinson, Jones, Rahman, & Rimes, 2018) found CGN was significantly correlated with depression at 18 years and with suicidal injury/self-harm at 16. However, these outcomes focused on late adolescence and this study did not report these correlations separately for males and females.

Retrospective studies have limitations as it is possible that participants may remember the gender typicality of their behavior incorrectly (Alanko et al., 2009). Recall by participants experiencing low mood may also be biased towards more negative memories (Blaney, 1986). Furthermore, it is possible that an unmeasured third variable is associated with both self-reported CGN and emotional or relationship problems. Therefore, studies using prospective measures of CGN and later emotional or relationship difficulties, adjusting for possible confounding variables, are needed.

There are a number of difficult experiences associated with CGN which may contribute to the development of mental health problems. Retrospective studies report that adults who recalled being more gender atypical in childhood were more likely to report feeling different from other children (D'Augelli , Grossman, & Starks, 2008). Recalled CGN has been found to be associated with more negative peer relationships in adolescence and young adulthood, including peer rejection (Landolt, Bartholomew, Saffrey, Oram, & Perlman, 2004) and victimization (e.g. Roberts et al., 2013). One study found that around half of the increased depression rates in gender nonconforming (versus conforming) young adults was explained by bullying and abuse victimization (Roberts et al., 2013). Cross-sectional studies have also found that greater self-perceived gender atypicality in adolescence is associated with lower concurrent self-worth (Smith & Leaper, 2006). The influence of these factors (feeling different, lack of peer acceptance, bullying, abuse, and low self-esteem) on the relationship between CGN and subsequent mental health problems requires research.

The present study

The present study investigated prospectively whether higher levels of gender nonconforming behaviors, rated by parents in infancy and children at 8 years, were associated with peer and emotional problems rated by parents and teachers when the child was 6-16 years old. To provide preliminary information about reasons for any significant relationships, analyses were repeated

adjusting for potential contributory factors which had been assessed several years before the final outcome point. It was hypothesized that:

1) Higher parent-rated and child-rated gender nonconforming behaviors in childhood would be associated with greater emotional and peer problems across childhood and adolescence, and that this relationship would be stronger for boys.
2) Adjusting for abuse, bullying, lack of peer acceptance, feeling different and low self-esteem would reduce the strength of these associations.

Furthermore, as multiple outcome points were included, between 6 – 16 years, interactions between gendered-behavior and time were investigated.

Methods

Sample

Data from The Avon Longitudinal Study of Parents and Children (ALSPAC), a longitudinal birth cohort study, were analyzed. Pregnant women resident in Avon, UK with expected dates of delivery between 1st April 1991 and 31st December 1992 were invited to take part in the study. 14,541 pregnant women took part, resulting in 14,676 fetuses, 14,062 live births and 13,988 children alive at 1 year. Further recruitment increased the total number of pregnancies to 15,247, resulting in 14,775 live births, and 14,701 children alive at 1 year with any data collected after age 7 (Boyd et al., 2013; Fraser et al., 2013). The study website contains details of all the data that is available through a fully searchable data dictionary and variable selection tool (http://www.bristol.ac.uk/alspac/researchers/our-data/).

Analysis investigating parent-rated gender-typed behaviors was based on the 11,192 (male *n* =5,808) participants (72% of the original cohort) who were alive at 1 year and whose carer had completed the Preschool Activities Inventory (PSAI) at one or more time point (the 'PSAI sample'). Analysis investigating child-rated gender-typed behaviors was based on the 7,049 (male *n*=3,481) participants (46% of the original cohort) who were alive at 1 year and had completed the Childhood Activities Inventory (the 'CAI sample'). There were 5,038 participants who were included in both samples (45% of the PSAI sample and 71% of the CAI sample).

To assess for attrition bias, those with missing and non-missing predictors were compared using independent samples t-tests, separately for the PSAI and CAI samples.

Ethical approval

Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee (see http://www.aslpac.bris.ac.uk) and King's College London College Research Ethics Committee (ref. PNM/14/15-67).

Measures

Parent-rated gender-typed behaviors

Gender-typed behaviors were rated by carers at 30, 42 and 57 months using the Pre-school Activities Inventory (PSAI; Golombok & Rust, 1993), which has test-retest reliability of .66 for girls and .62 for boys. Parents rate how frequently their child displays 24 different behaviors, from 1 (never) to 5 (very often). Scores were standardized in line with the test instructions, with higher scores reflecting more masculine behavior and lower scores reflecting more feminine behavior. A mean score was calculated across the three time-points after imputation as preliminary analysis showed that the scores on the three measures were highly correlated with one another and their relationships with the outcomes were similar. Separate standardized Z-scores were calculated for males and females. Some items from the 'likes and dislikes' subscale overlap in content with items from the 'emotional problems' subscale of the SDQ (e.g. nervous in new situations); therefore, PSAI scores were also calculated without the items from this subscale. However, findings were highly similar when using this modified score, therefore the original PSAI score was used.

Child-rated gender-typed behaviors

Gender-typed behaviors were self-rated by children at 8.5 years using the Childhood Activities Inventory (CAI), an adapted and shortened 16-item form of the PSAI which has split-half reliability of .64 in both boys and girls (Golombok et al., 2008). Separate standardized Z-scores were calculated for males and females.

Emotional and peer problems

Emotional and peer difficulties were assessed using the 'emotional problems' and 'peer problems' subscales of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997). The SDQ has internal consistency of α=.67 and .78 for emotional problems, and .57 and .70 for peer problems, for parent and teacher-ratings respectively (Goodman, 2001). Ratings were made by parents at 6, 9, 11, 13 and 16 years, and by teachers at 7 and 10 years. As bullying was investigated as a covariate, a peer problems score was also calculated without the item "picked on or bullied by other children". However, analysis demonstrated that using this modified score did not change the results, therefore the full peer problems score was used.

Psychosocial covariates

Self-esteem

Self-esteem was rated by children at 8.5 years using the 'global self-worth' subscale of Harter's Self Perception Profile for Children, which has good internal consistency (α=.78-.87; Harter, 2012).

Bullying victimization

Children self-reported on their experience of peer victimization at 8.5, 10.5 and 12.5 years using a shortened version of the Bullying and Friendship Interview Schedule (Wolke, Woods, Bloomfield, & Karstadt, 2000). Children were asked whether certain events had happened in the past 6 months and how frequently. Events were separated into 'overt' and 'relational' bullying. Following author recommendations, children were classed as victims of overt bullying if they had experienced any of the five overt events frequently (several times a month) or very frequently (several times a week)

and were classed as victims of relational bullying if they had experienced any of the four relational events frequently or very frequently (Wolke et al., 2000). Two binary variables were generated to indicate whether children had experienced overt or relational bullying ever (at any of the 3 timepoints; Lereya, Copeland, Costello, & Wolke, 2015).

Abuse

Children's experiences of being physically hurt ('child was physically hurt by someone') and sexually abused ('child was sexually abused by someone') were reported by mothers at 1, 2, 4, 5, 6 and 8 years. Mothers also reported on children's experiences of emotional (at 1, 3, 5 and 6 years) and physical (at 8 months and 2, 9 and 11 years; 'you/your partner were emotionally/physically cruel to your child') cruelty. All responses were yes/no. A binary variable was generated to indicate whether a child had at least one experience of any type of abuse at any of the time points (Lereya et al., 2015). A binary rather than a count variable was used as some responses may refer to the same experience.

Feeling accepted and Feeling different

At 11 years, children rated 'my school is a place where other people accept me as I am' (agree/mostly agree/mostly disagree/disagree). At 13 years, children answered 'how often do you feel different from others?' (always/mostly/sometimes/not often/never). Due to similarity of these items to some within the peer problems SDQ scale, they were only investigated as covariates for the emotional problems outcomes.

Demographic covariates

Demographic covariates related to the predictors and/or outcomes were controlled for. Mothers reported their age at 18 weeks gestation, and their highest educational level (below O level v O level v A level or above) and their child's ethnicity (white v non-white) at 32 weeks gestation. Mothers completed the Edinburgh Postnatal Depression Scale (Cox , Holden, & Sagovsky, 1987) at 32 weeks gestation, and when the child was 8 weeks and 2 years 9 months old. Children's IQ at 8.5 years was measured using the WISC-III (Wechsler, Golombok, & Rust, 1992). To control for the possible influence of siblings on gendered behavior, the presence of brothers and sisters at 6 years was also controlled for.

Data analysis

Missing Values

Due to its longitudinal nature, there are high levels of missing data in the ALSPAC dataset (see Appendix 1). Using complete-case analysis can bias results (White, Royston, & Wood, 2011). Multiple imputation is a statistical technique used to overcome this issue, which creates multiple data sets where the missing values are replaced with plausible values estimated in a principled way using the observed data values. Each data set is analyzed separately providing the estimate of interest, then estimates are combined using Rubin's rules (Rubin, 1987). The 'mi impute chained' command in Stata 15.0 was used, as this allows for missing values across several variables and assumes an arbitrary missing data pattern. Separate imputations were carried out for the sample with complete CAI predictor, and for the sample with at least one PSAI predictor. Each of these imputation commands were run separately by sex, as the direction of scoring of the PSAI and CAI depend on sex. Following recommendations based on the amount of missing data (White et al., 2011), 100 imputations were carried out in each case. Auxiliary variables were included which predicted the incomplete variables (see Appendix 1).

Statistical analysis

For each sex separately, 8 individual linear mixed regression models with random intercepts were fitted to the repeated measures SDQ outcome: standardized parent-rated and child-rated gender-typed behavior were analyzed separately as predictors of each of the four different outcomes (parent/teacher-rated emotional/peer problems), adjusting for covariates. For the teacher-rated outcome models, robust standard errors were used to address heteroskedasticity. Models were fitted using Stata version 15.0 (StataCorp LLC, College Station, TX)

Initial unadjusted models were fitted on the complete case data to examine the relationship between each predictor and outcome (see Appendix 2). Subsequent analyses were fitted on the imputed data (Tables 1 and 2). Initial unadjusted models examined the relationship between each predictor and outcome (Model 1). These results were similar to those reported for the complete-case data.

If gender-typed behavior was a significant predictor of the outcome, models were then adjusted for demographic covariates (Model 2), including those which may be related to attrition bias (e.g. child's ethnicity, mother's age). If gender-typed behavior was still a significant predictor of the outcome, models were then adjusted for each of the psychosocial covariates individually. If the addition of any covariate reduced the beta coefficient for gender-typed behavior by 5% or more it was added to Model 3. Model 3 consisted of the predictor, the demographic confounders, and any of the psychosocial covariates (simultaneously) which had reduced the beta coefficient of the predictor by 5% or more. Model 3 was only fitted if any psychosocial covariates were found to reduce the beta coefficient for gender-typed behavior by 5% or more (see Appendix 3 for individual results of covariates included).

The models were checked for time by gender-typed behavior interactions by including these in the final models, however there was no strong evidence for a difference in effects over time, therefore the results reported are from models without interaction terms.

Results

Sample characteristics

For sample characteristics, see Appendix 1. In the PSAI sample, before imputation, mean PSAI scores were 62.4 (*SD* = 7.4) for males and 37.6 (*SD* = 7.8) for females. Most participants (89.4%) were white and most mothers (35.9%) were educated to A level or above. In the CAI sample, mean CAI scores were 59.9 (*SD* = 11.4) for males and 40.0 (*SD* = 12.41) for females. Most participants (87.4%) were white and most mothers (40.2%) were educated to A level or above.

Attrition Bias

In comparison to the PSAI sample (*n*=11,192), the non-PSAI sample (those with no data for any of the PSAI measures; *n*=3,492) had higher rates of parent-rated emotional problems at 6, 9, 11 and 13 years ($p < .05$); parent-rated peer problems at 9 and 11 years ($p < .03$); teacher-rated emotional and peer problems ($p < .001$); younger mothers ($p < .001$); mothers with higher postnatal depression ($p < .001$); non-white participants ($p < .001$); mothers with the lowest level of education ($p < .001$); and children who had experienced abuse ($p < .001$); and relational bullying ($p < .001$).

In comparison to the CAI sample (*n*=7,049), the non-CAI sample (those who did not have data for the CAI measure; *n*=7,635) had higher rates of parent-rated emotional problems at 9, 11, 13 and 16 years ($p < .003$); parent-rated peer problems at all time points ($p < .001$); teacher-rated emotional and peer problems ($p < .001$); younger mothers ($p < .001$); mothers with higher postnatal depression ($p < .001$); non-white participants ($p < .001$); mothers with below O level education ($p < .001$); children who had experienced abuse ($p < .001$); overt bullying ($p < .001$); and relational bullying ($p < .001$); and children who felt less accepted by peers ($p = .004$).

Linear mixed model analyses: Parent-rated gender-typed behavior

Emotional Problems

In the unadjusted models (Model 1, Table 1) for boys, lower (more feminine) PSAI scores were significantly associated with greater parent-rated and teacher-rated emotional problems even after adjustment for demographic confounders (Model 2). Adding 'feeling different' or 'feeling accepted' reduced the beta coefficient for PSAI for both models (by 5.0% and 5.7% respectively for parent-report; 9.7% and 7.8% for teacher-report). In Model 3, PSAI scores remained significantly associated with parent and teacher-rated emotional problems.

In the unadjusted models for girls, lower (more feminine) PSAI scores were significantly associated with greater parent-rated emotional problems (Model 1, Table 2) even after adjustment for demographic confounders (Model 2). The beta value for PSAI was not reduced by 5% or more by adding any of the covariates of interest, therefore Model 3 was not fitted in this case. PSAI scores were not significantly associated with teacher-rated emotional problems for girls (Model 1, Table 2).

Peer Problems

In the unadjusted models for boys, lower PSAI scores were significantly associated with greater parent-rated and teacher-rated peer problems (Model 1, Table 1), even after adjustment for demographic confounders (Model 2). The beta value for PSAI was not reduced by 5% or more in either model by adding any other covariates.

In the unadjusted models for girls, *higher* (more masculine) PSAI scores were significantly associated with greater parent-rated and teacher-rated peer problems (Model 1, Table 2), even after adjustment for demographic confounders (Model 2). Adding self-esteem, abuse, overt or relational bullying victimization reduced the beta coefficients for PSAI in the parent-rated outcome model (by 9.0%, 8.6%, 13.7% and 5.0%, respectively). For the teacher-rated peer problems model, adding abuse or overt bullying victimization reduced the beta coefficients for PSAI (by 5.6% and 8.5%,

respectively). In both subsequent models (Model 3), higher PSAI scores remained significantly associated with peer problems for girls.

Linear mixed model analyses: Child-rated gender-typed behavior

Emotional Problems

In the unadjusted models for boys, lower (more feminine) CAI score was significantly associated with greater parent-rated and teacher-rated emotional problems (Model 1, Table 1), even after adjustment for demographic confounders (Model 2). Adding self-esteem, relational bullying victimization, feeling different or feeling accepted reduced the beta coefficient for CAI in the parent-rated outcome model (by 5.2%, 8.6%, 12.5% and 15.3%, respectively). For the teacher-rated model, adding self-esteem score or feeling different reduced the beta coefficient for CAI (by 5.7% and 14.2%, respectively). In both subsequent models (Model 3), CAI remained significantly associated with emotional problems for boys.

In the unadjusted models for girls, lower (more feminine) CAI scores were significantly associated with greater parent-rated and teacher-rated emotional problems (Model 1, Table 2), even after adjustment for demographic confounders (Model 2). For the model with parent-rated emotional problems, adding relational bullying victimization reduced the beta coefficient for CAI by 9.6%. In the subsequent model (Model 3), CAI remained significantly associated with emotional problems. For teacher-rated emotional problems, the beta value for CAI was not reduced by 5% or more by adding any of the covariates of interest.

Peer Problems

In the unadjusted models for boys, lower CAI scores were significantly associated with greater parent-rated and teacher-rated peer problems (Model 1, Table 1), even after adjustment for

demographic confounders (Model 2). Adding self-esteem score or relational bullying victimization reduced the beta coefficient for CAI for both the parent-rated (by 7.1% and 7.3% respectively) and teacher-rated peer problems (by 8.4% and 7.0% respectively). In both subsequent models (Model 3), CAI remained significantly associated with peer problems for boys.

In the unadjusted models for girls, lower CAI scores were significantly associated with greater parent-rated peer problems (Model 1, Table 2), even when adjusting for demographic confounders (Model 2). Adding relational bullying victimization reduced the beta coefficient for CAI by 8.6%. In Model 3, lower CAI remained significantly associated with more peer problems. CAI scores were not significantly associated with teacher-rated peer problems for girls (Model 1, Table 2).

[Table 1]

[Table 2]

Discussion

Childhood gender nonconformity and emotional problems

This study demonstrated for the first time that there is a prospective relationship between gender-nonconforming behavior and emotional problems in childhood and adolescence, both when behavior is rated by carers in infancy and by self-report at 8 years. For boys, more gender-nonconforming behavior rated by their carers and themselves was associated with more parent and teacher-rated emotional problems. In contrast, for girls, greater parent-rated or self-reported gender-nonconforming behavior was associated with *fewer* emotional problems when rated by parents, and the same pattern was found for self-reported gender-nonconforming behavior and teacher-rated emotional functioning. There was no evidence of a significant interaction between time (between ages of 6 to 16 years) and gender-typed behavior.

Some previous studies have only found an association between CGN and emotional problems for males and not females (e.g. Lippa, 2008), but the authors are not aware of any that found results in the opposite direction for females. A study in adolescents found self-rated increases in masculinity over time are associated with fewer depressive symptoms in both males and females (e.g. Priess, Lindberg, & Hyde, 2009). However, associations between femininity and depressive symptoms have been weaker (e.g. Stoppard & Paisley, 1987) or not found (e.g. Priess et al., 2009).

One explanation for the association between gender-nonconforming behavior and more emotional problems for boys only, is that the strength and nature of gendered expectations may differ for the sexes, e.g. gender-nonconformity may be less accepted in boys than girls, both by other children (e.g. Conry-Murray, Kim, & Turiel, 2015) and by parents (Spivey, Huebner, & Diamond, 2018). Stigma processes may then contribute to poorer mental health. Indeed, when models were adjusted for feeling accepted by peers, the association between gender-nonconforming behavior and worse emotional problems in boys was generally weakened. Similarly, adjusting for relational bullying reduced the association between child-rated CGN and parent-rated emotional problems. Adjusting

for feeling different and lower self-esteem, which may reflect internalized stigma, also reduced the association between child-rated gendered behavior and emotional problems for boys.

It is also possible that the association between gender-nonconforming behavior and emotional problems for boys is related to aspects of 'femininity' or lower 'masculinity', rather than CGN. This would be consistent with the finding that both boys and girls rated as more feminine experienced worse emotional problems. Gender differences in emotional problems are often conceptualized as being due to biological factors such as sex hormones (e.g.Li & Graham, 2017), and there is some evidence of an association between exposure to prenatal sex hormones and differences in gender-typed play behavior in childhood (Auyeung et al., 2009). There are different mechanisms through which childhood behaviors may be associated with subsequent emotional problems. For example, rough-and-tumble and exercise play in pre-school children, usually considered 'male-typical' activities, have been found to be associated with improved emotion regulation one year later (Lindsey & Colwell, 2013). Children showing more female-typical behaviors may also be more at risk of bullying by being less physically active. Future research should investigate gene-environment interactions and correlations in understanding the relationship between gendered behavior and subsequent emotional or peer problems.

The relationship between childhood gender-typed behaviors and distress is likely to be influenced by multiple social factors including parenting. For example, parents are more likely to encourage emotion expression (Cox, Mezulis, & Hyde, 2010) in daughters than sons. Dwelling on negative feelings, or rumination, is associated with the increased risk for depression found in women (Jose & Brown, 2008). Future research should investigate the role of other aspects of parenting including gendered expectations and responses to different forms of gendered behavior.

Childhood gender nonconformity and peer problems

For both boys and girls, parent-rated gender-nonconforming behaviors were associated with more parent and teacher-rated peer problems. However, for girls, *self*-rated gender-nonconforming behaviors were associated with *fewer* parent-rated peer problems, with no significant association for teacher-rated peer problems. The parent ratings of gender-typed behavior are made much earlier than the child ratings, so it is possible these findings reflect age effects on peer relational associations with gender-typed behavior or within-individual age-related changes in gender-typed behavior. Child ratings were also taken in closer time proximity to the measures of peer functioning so may be more accurate ratings of current behaviors. Although parent and child ratings in this cohort are significantly associated (Golombok et al., 2008), both raters have access to different types of information, e.g. parents can only rate behaviors they observe or hear about. Similarly, teachers are rating peer relationships at school whereas parents are more likely to observe peer relationships with chosen friends outside of school. It is possible that for girls, gender-nonconforming behavior is associated with more peer difficulties at school but fewer peer difficulties with friends outside of school.

Adjusting for bullying or abuse victimization weakened associations between gender-nonconforming behavior and greater peer problems. Bullying and abuse are associated with lower self-esteem (e.g. Gruber & Fineran, 2007) which may contribute to peer difficulties. Adjusting for self-esteem reduced the strength of many of the associations between self-rated gender-nonconforming behavior and greater peer problems. The impact of bullying, abuse victimization and lower self-esteem on peer relationships in gender nonconforming children requires further investigation.

Strengths and limitations

The study's prospective design addresses the risk of bias associated with retrospective recollections of CGN. Findings are also strengthened by the inclusion of both observer and self-ratings of gender-typed behavior, and parent and teacher outcome ratings.

Methodological limitations include modest internal consistency for the outcome measures and no child self-reports of outcomes. Additionally, many measures were only available at certain ages and therefore mediational analyses were not undertaken, as for such analyses hypothesized mediators should be assessed at an intermediate time point between the predictor and outcome.

Implications

Further research is needed into whether female-typical behaviors are an indirect marker of risk for mental illness (e.g. via biological or social factors), or whether they play a contributory role. Those working with children should be aware of the possible effects of behaviors they reinforce in boys and girls. Anti-bullying strategies should be stringently enforced for all children. As always, clinicians working with distressed children and adolescents should assess for experiences of bullying, abuse victimization, feeling different and lower self-esteem that may be contributing to emotional problems.

Conclusion

This is the first prospective study to demonstrate that more female-typical behaviors in boys have a small but significant association with greater subsequent emotional and peer problems in childhood and adolescence. Possible contributory factors for future research to investigate include feeling unaccepted, feeling different, relational bullying victimization and lower self-esteem. However, for girls, more female-typical behaviors were similarly associated with increased risk for emotional problems, and also for peer problems when child-ratings of behaviors were used. Future research should investigate whether activities typically reinforced for girls may contribute to their increased risk for emotional problems or whether they may be a marker for another type of risk factor.

Acknowledgements

The authors are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and they will serve as guarantors for the contents of this paper. AW is supported by a studentship from the ESRC LISS DTP. KG is in part supported by the NIHR Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust and Institute of Psychiatry, Psychology & Neuroscience, King's College London. This report is independent research partly arising from a Doctoral Research Fellowship supported by the National Institute for Health Research. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.The authors have declared that they have no competing or potential conflicts of interest.

Key Points

- Childhood gender nonconformity has been found to be associated with more emotional and peer problems, but previous research has been cross-sectional or retrospective.
- Here, for boys, higher parent and child ratings of female-typical behavior taken in childhood were associated with greater subsequent emotional and peer problems in childhood and adolescence. Association sizes were reduced when adjusting for self-esteem and interpersonal factors such as being bullied.
- For girls, higher self-rated female-typical behaviors (i.e. *fewer* gender-nonconforming activities) were associated with more emotional and peer problems.

- Research is needed into why female-typical behaviors are associated with greater subsequent emotional and peer problems for both boys and girls.

Correspondence

Katharine A. Rimes

Dept. of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London SE5 8AF, UK

Email: Katharine.Rimes@kcl.ac.uk

References

Alanko, K., Santtila, P., Witting, K., Varjonen, M., Jern, P., Johansson, A., . . . Kenneth Sandnabba, N. (2009). Psychiatric symptoms and same-sex sexual attraction and behavior in light of childhood gender atypical behavior and parental relationships. *Journal of Sex Research, 46*(5), 494-504. doi:10.1080/00224490902846487

Auyeung, B., Baron-Cohen, S., Ashwin, E., Knickmeyer, R., Taylor, K., Hackett, G., & Hines, M. (2009). Fetal testosterone predicts sexually differentiated childhood behavior in girls and in boys. *Psychological Science, 20*(2), 144-148. doi:10.1111/j.1467-9280.2009.02279.x

Blaney, P. H. (1986). Affect and memory: a review. *Psychol Bull, 99*(2), 229-246.

Boyd, A., Golding, J., Macleod, J., Lawlor, D. A., Fraser, A., Henderson, J., . . . Davey Smith, G. (2013). Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *International Journal of Epidemiology, 42*(1), 111-127. doi:10.1093/ije/dys064

Conry-Murray, C., Kim, J. M., & Turiel, E. (2015). Judgments of gender norm violations in children from the United States and Korea. *Cognitive Development, 35*, 122-136. doi:https://doi.org/10.1016/j.cogdev.2015.04.002

Cox , J. L., Holden, J. M., & Sagovsky, R. (1987). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *British Journal of Psychiatry, 150*, 782-786.

Cox, S. J., Mezulis, A. H., & Hyde, J. S. (2010). The influence of child gender role and maternal feedback to child stress on the emergence of the gender difference in depressive rumination in adolescence. *Developmental Psychology, 46*(4), 842-852. doi:10.1037/a0019813

D'Augelli , A. R., Grossman, A. H., & Starks, M. T. (2008). Gender Atypicality and Sexual Orientation Development Among Lesbian, Gay, and Bisexual Youth. *Journal of Gay & Lesbian Mental Health, 12*(1-2), 121-143. doi:10.1300/J529v12n01_08

Fraser, A., Macdonald-Wallis, C., Tilling, K., Boyd, A., Golding, J., Davey Smith, G., . . . Lawlor, D. A. (2013). Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *International Journal of Epidemiology, 42*(1), 97-110. doi:10.1093/ije/dys066

Friedman, M. S., Koeske, G. F., Silvestre, A. J., Korr, W. S., & Sites, E. W. (2006). The impact of gender-role nonconforming behavior, bullying, and social support on suicidality among gay male youth. *Journal of Adolescent Health, 38*(5), 621-623. doi:10.1016/j.jadohealth.2005.04.014

Golombok, S., & Rust, J. (1993). The Pre-School Activities Inventory: A standardized assessment of gender role in children. *Psychological Assessment, 5*(2), 131-136. doi:10.1037/1040-3590.5.2.131

Golombok , S., Rust, J., Zervoulis, K., Croudace, T., Golding, J., & Hines, M. (2008). Developmental trajectories of sex-typed behavior in boys and girls: A longitudinal general population study of children aged 2.5-8 years. *Child Development, 79*(5), 1583-1593. doi:DOI 10.1111/j.1467-8624.2008.01207.x

Golombok , S., Rust, J., Zervoulis, K., Golding, J., & Hines, M. (2012). Continuity in sex-typed behavior from preschool to adolescence: a longitudinal population study of boys and girls aged 3-13 years. *Arch Sex Behav, 41*(3), 591-597. doi:10.1007/s10508-011-9784-7

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38*(5), 581-586. doi:DOI 10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1337-1345. doi:10.1097/00004583-200111000-00015

Gruber, J. E., & Fineran, S. (2007). The Impact of Bullying and Sexual Harassment on Middle and High School Girls. *Violence Against Women, 13*(6), 627-643. doi:10.1177/1077801207301557

Harter, S. (2012). *Self-Perception Profile for Children: Manual and Questionnaires (Grades 3-8)*: University of Denver.

Jones, A., Robinson, E., Oginni, O., Rahman, Q., & Rimes, K. A. (2017). Anxiety disorders, gender nonconformity, bullying and self-esteem in sexual minority adolescents: Prospective birth cohort study. *Journal of Child Psychology and Psychiatry, 58*(11), 1201-1209.

Jose, P. E., & Brown, I. (2008). When does the Gender Difference in Rumination Begin? Gender and Age Differences in the Use of Rumination by Adolescents. *Journal of Youth and Adolescence, 37*(2), 180-192. doi:10.1007/s10964-006-9166-y

Landolt, M. A., Bartholomew, K., Saffrey, C., Oram, D., & Perlman, D. (2004). Gender nonconformity, childhood rejection, and adult attachment: a study of gay men. *Archives of Sexual Behavior, 33*(2), 117-128.

Lereya, S. T., Copeland, W. E., Costello, E. J., & Wolke, D. (2015). Adult mental health consequences of peer bullying and maltreatment in childhood: two cohorts in two countries. *The Lancet Psychiatry, 2*(6), 524-531. doi:10.1016/S2215-0366(15)00165-0

Li, S. H., & Graham, B. M. (2017). Why are women so vulnerable to anxiety, trauma-related and stress-related disorders? The potential role of sex hormones. *The Lancet Psychiatry, 4*(1), 73-82. doi:10.1016/s2215-0366(16)30358-3

Lindsey, E. W., & Colwell, M. J. (2013). Pretend and Physical Play: Links to Preschoolers' Affective Social Competence. *Merrill-Palmer Quarterly, 59*(3), Article 4.

Lippa, R. A. (2008). The Relation Between Childhood Gender Nonconformity and Adult Masculinity–Femininity and Anxiety in Heterosexual and Homosexual Men and Women. *Sex Roles, 59*(9-10), 684-693. doi:10.1007/s11199-008-9476-5

Oginni, O. A., Robinson, E. J., Jones, A., Rahman, Q., & Rimes, K. A. (2018). Mediators of increased self-harm and suicidal ideation in sexual minority youth: a longitudinal study. *Psychol Med*, 1-9. doi:10.1017/s003329171800346x

Priess, H. A., Lindberg, S. M., & Hyde, J. S. (2009). Adolescent gender-role identity and mental health: gender intensification revisited. *Child Development, 80*(5), 1531-1544. doi:10.1111/j.1467-8624.2009.01349.x

Rieger, G., & Savin-Williams, R. C. (2012). Gender nonconformity, sexual orientation, and psychological well-being. *Archives of Sexual Behavior, 41*(3), 611-621. doi:10.1007/s10508-011-9738-0

Roberts, A. L., Rosario, M., Slopen, N., Calzo, J. P., & Austin, S. B. (2013). Childhood gender nonconformity, bullying victimization, and depressive symptoms across adolescence and early adulthood: an 11-year longitudinal study. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(2), 143-152. doi:10.1016/j.jaac.2012.11.006

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Smith, T. E., & Leaper, C. (2006). Self-Perceived Gender Typicality and the Peer Context During Adolescence. *Journal of Research on Adolescence, 16*(1), 91-104. doi:doi:10.1111/j.1532-7795.2006.00123.x

Spivey, L. A., Huebner, D. M., & Diamond, L. M. (2018). Parent responses to childhood gender nonconformity: Effects of parent and child characteristics. *Psychology of Sexual Orientation and Gender Diversity*, No Pagination Specified.

Stoppard, J. M., & Paisley, K. J. (1987). Masculinity, femininity, life stress, and depression. *Sex Roles, 16*(9), 489-496. doi:10.1007/bf00292483

Wechsler, D., Golombok, S., & Rust, J. (1992). *WISC-III UK Wechsler Intelligence Scale for Children – Third Edition UK Manual*. Sidcup, UK: The Psychological Corporation.

White, I. R., Royston, P., & Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine, 30*(4), 377-399. doi:10.1002/sim.4067

Wolke, D., Woods, S., Bloomfield, L., & Karstadt, L. (2000). The association between direct and relational bullying and behaviour problems among primary school children. *Journal of Child Psychology and Psychiatry, .41*(8), pp. doi:10.1111/1469-7610.00687 11099116

Table 1 Results of post-imputation linear mixed analyses for boys with masculine-typed behavior as a predictor.

		Emotional Problems				Peer Problems			
		Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated
		Parent-rated gender-typed behavior (standardized PSAI score)		Child-rated gender-typed behavior (standardized CAI score)		Parent-rated gender-typed behavior (standardized PSAI score)		Child-rated gender-typed behavior (standardized CAI score)	
Model 1	*β*	-.122	-.071	-.083	-.101	-.155	-.114	-.122	-.124
	95% CI	-.160, -.084	-.112, -.023	-.125, -.040	-.163, -.039	-.191, -.119	-.165, -.064	-.163, -.080	-.186, -.061
	p	<.001	.004	<.001	.002	<.001	<.001	<.001	<.001
Model 2	*β*	-.129	-.089	-.075	-.085	-.165	-.132	-.119	-.119
	95% CI	-.167, -.092	-.139, -.039	-.115, -.034	-.147, -.023	-.202, -.128	-.184, -.080	-.160, -.078	-.181, -.056
	p	<.001	<.001	<.001	.007	<.001	<.001	<.001	<.001
Model 3	*β*	-.118	-.076	-.052	-.070	[b]	[b]	-.103	-.102
	95% CI	-.155, -.081	-.126, -.026	-.093, -.012	-.132, -.007			-.144, -.0624	-.164, -.040
	p	<.001	.003	.011	.029			<.001	.001

Higher PSAI/CAI scores indicate more masculine behavior.

Model 1: Unadjusted. Model 2: Adjusted for demographic confounders. Model 3: Further adjusted for any covariates which reduced *β* in Step 2 by 5% or more.

[a] No further analysis performed as gender-typed behavior not a significant predictor of the outcome.

[b] No further analysis performed as no covariates reduced the beta coefficient of gender-typed behavior by 5% or more.

Table 2 Results of post-imputation linear mixed analyses for girls with masculine-typed behavior as a predictor.

		Emotional Problems				Peer Problems			
		Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated
		Parent-rated gender-typed behavior (standardized PSAI score)		Child-rated gender-typed behavior (standardized CAI score)		Parent-rated gender-typed behavior (standardized PSAI score)		Child-rated gender-typed behavior (standardized CAI score)	
Model 1	*β*	-.061	.004	-.068	-.121	.074	.144	-.073	-.032
	95% CI	-.103, -.019	-.052, .059	-.116, -.021	-.175, -.066	.043, .106	.101, .188	-.108, -.038	-.080, .017
	P	.005	.902	.005	<.001	<.001	<.001	<.001	.202
Model 2	*β*	-.064	[a]	-.054	-.099	.085	.147	-.058	[a]
	95% CI	-.105, -.023		-.100, -.009	-.152, -.046	.054, .117	.104, .190	-.092, -.024	
	P	.002		.020	<.001	<.001	<.001	.001	
Model 3	*β*	[b]		-.049	[b]	.062	.127	-.053	
	95% CI			-.095, -.004		.031, .093	.084, .170	-.087, -.019	
	p			.035		<.001	<.001	.002	

Higher PSAI/CAI scores indicate more masculine behavior.

Model 1: Unadjusted. Model 2: Adjusted for demographic confounders. Model 3: Further adjusted for any covariates which reduced *β* in Step 2 by 5% or more.

[a] No further analysis performed as gender-typed behavior not a significant predictor of the outcome.

[b] No further analysis performed as no covariates reduced the beta coefficient of gender-typed behavior by 5% or more.

Appendix 1 Sample characteristics and percentage of missing data for each group before imputation.

	Parent-rated gender-typed behaviour sample		Child-rated gender-typed behaviour sample	
	Males (n=5,808)	Females (n=5,384)	Males (n=3,481)	Females (n=3,568)
Gender-typed behaviour				
PSAI score, *M (SD)*				
30 months	60.8 (8.3)	40.8 (8.5)		
% missing (n)	*13.5 (786)*	*14.3 (772)*		
42 months	62.4 (8.6)	37.0 (9.3)		
% missing (n)	*10.5 (610)*	*10.2 (550)*		
57 months	64.3 (8.8)	35.1 (9.4)		
% missing (n)	*19.3 (1,119)*	*18.9 (1,020)*		
CAI Score, *M (SD)*			59.9 (11.4)	40.0 (12.4)
% missing (n)			*0 (0)*	*0 (0)*
Demographic confounders				
Mother				
Age (years), *M (SD)*	28.2 (4.8)	28.0 (4.7)	28.9 (4.7)	28.7 (4.5)
% missing (n)	*2.8 (164)*	*2.7 (143)*	*5.9 (207)*	*6.8 (242)*
Highest education, *% (n)*				
Below O Level	26.1 (1,515)	25.6 (1,379)	20.4 (709)	19.7 (704)
O Level	34.0 (1,975)	33.6 (1,808)	32.9 (1,144)	31.6 (1,127)
A Level or above	35.4 (2,058)	36.5 (1,965)	39.5 (1,376)	40.8 (1,455)
Missing	*4.5 (260)*	*4.3 (232)*	*7.2 (252)*	*7.9 (282)*
Postnatal depression score				
32 weeks gestation, *M (SD)*	6.9 (5.0)	6.9 (5.0)	6.6 (4.9)	6.3 (4.8)
% missing (n)	*7.6 (442)*	*7.0 (378)*	*10.0 (350)*	*10.1 (360)*
8 weeks, *M (SD)*	6.0 (4.7)	5.9 (4.7)	5.8 (4.5)	5.8 (4.6)

% missing (n)	*7.8 (443)*	*6.9 (373)*	*9.9 (343)*	*9.9 (352)*
2 years 9 months, *M (SD)*	6.3 (5.0)	6.2 (5.1)	6.1 (4.8)	6.0 (4.9)
% missing (n)	*15.9 (443)*	*15.0 (806)*	*16.5 (573)*	*17.3 (618)*
Child				
Ethnicity, *% (n)*				
White	89.2 (5,178)	89.8 (4,833)	87.3 (3,041)	87.4 (3,120)
Non-white	4.2 (243)	3.9 (211)	3.5 (123)	3.4 (122)
Missing	*6.7 (387)*	*6.3 (340)*	*9.1 (317)*	*9.1 (326)*
IQ, *M (SD)*	104.5 (17.0)	104.3 (15.9)	104.2 (17.0)	104.1 (15.9)
% missing (n)	*42.3 (2,454)*	*37.5 (2,018)*	*1.7 (58)*	*1.2 (44)*
Presence of brothers, *% (n)*				
Yes	40.5 (2,354)	41.7 (2,247)	45.3 (1,579)	44.5 (1,589)
No	33.9 (1,970)	34.1 (1,834)	37.2 (1,296)	37.2 (1,329)
Missing	*25.6 (1,484)*	*24.2 (1,303)*	*17.4 (606)*	*18.2 (650)*
Presence of sisters, *% (n)*				
Yes	40.2 (2,337)	40.4 (2,176)	44.3 (1,542)	37.6 (1,577)
No	34.2 (1,987)	35.4 (1,905)	38.3 (1,333)	65.6 (1,341)
Missing	*25.6 (1,484)*	*24.2 (1,303)*	*17.4 (606)*	*18.2 (650)*
Outcomes				
Emotional Problems, *M (SD)*				
6 years 9 months	1.4 (1.7)	1.6 (1.7)	1.4 (1.6)	1.6 (1.7)
% missing (n)	*26.2 (1,523)*	*24.9 (1,342)*	*17.9 (624)*	*18.8 (669)*
7 years	1.3 (1.9)	1.4 (1.9)	1.2 (1.8)	1.3 (1.9)
% missing (n)	*54.0 (3,137)*	*52.6 (2,830)*	*52.7 (1,833)*	*52.3 (1,867)*
9 years 10 months	1.4 (1.7)	1.6 (1.8)	1.3 (1.7)	1.6 (1.8)
% missing (n)	*34.5 (2,002)*	*31.4 (1,688)*	*16.1 (560)*	*15.7 (560)*
10 years	1.3 (1.8)	1.3 (1.9)	1.2 (1.8)	1.2 (1.8)
% missing (n)	*47.9 (2,781)*	*46.3 (2,492)*	*44.8 (1,561)*	*42.7 (3,568)*
11 years 8 months	1.3 (1.7)	1.6 (1.8)	1.3 (1.6)	1.6 (1.8)

% missing (n)	*41.1 (2,389)*	*36.7 (1,976)*	*23.0 (802)*	*21.6 (772)*
13 years 1 month	1.2 (1.6)	1.6 (1.8)	1.8 (1.5)	1.6 (1.7)
% missing (n)	*43.3 (2,514)*	*39.1 (2,107)*	*26.0 (904)*	*24.6 (877)*
16 years 6 months	1.1 (1.56)	1.9 (2.0)	1.1 (1.5)	1.8 (2.0)
% missing (n)	*55.5 (3,224)*	*49.3 (2,657)*	*40.0 (1,391)*	*35.2 (1,255)*
Peer Problems, *M (SD)*				
6 years 9 months	1.2 (1.5)	1.0 (1.3)	1.1 (1.5)	0.9 (1.3)
% missing (n)	*26.2 (1,524)*	*24.8 (1,337)*	*18.0 (627)*	*18.7 (667)*
7 years	1.4 (1.9)	1.0 (1.6)	1.2 (1.8)	0.9 (1.5)
% missing (n)	*54.0 (3,137)*	*52.6 (2,830)*	*52.7 (1,833)*	*52.3 (1,867)*
9 years 10 months	1.2 (1.6)	1.0 (1.4)	1.1 (1.5)	1.0 (1.4)
% missing (n)	*34.3 (1,992)*	*31.4 (1,692)*	*16.0 (557)*	*15.8 (564)*
10 years	1.4 (2.0)	1.0 (1.6)	1.3 (1.9)	0.9 (1.6)
% missing (n)	*47.9 (2,780)*	*46.2 (2,492)*	*44.8 (1,561)*	*42.7 (1,522)*
11 years 8 months	1.2 (1.6)	1.0 (1.5)	1.1 (1.6)	1.0 (1.4)
% missing (n)	*40.9 (2,376)*	*36.6 (1,969)*	*22.8 (794)*	*21.6 (769)*
13 years 1 month	1.3 (1.7)	1.1 (1.5)	1.3 (1.7)	1.0 (1.4)
% missing (n)	*43.3 (2,517)*	*39.1 (2,106)*	*26.0 (906)*	*24.6 (877)*
16 years 6 months	1.2 (1.5)	1.0 (1.4)	1.2 (1.5)	1.0 (1.4)
% missing (n)	*55.5 (3,225)*	*49.3 (2,657)*	*40.0 (1,390)*	*35.2 (1,256)*
Covariates of interest				
Abuse victimisation, *% (n)*				
Yes	20.1 (1,167)	16.7 (898)	20.9 (727)	16.3 (581)
No	30.5 (1,769)	34.2 (1,844)	39.8 (1,386)	42.8 (1,526)
Missing	*49.4 (2,872)*	*45.7 (2,642)*	*39.3 (1,368)*	*40.9 (1,461)*
Overt bullying victim, *% (n)*				
Yes	26.7 (1,548)	26.6 (1,433)	38.8 (1,351)	37.7 (1,344)
No	22.4 (1,301)	27.0 (1,456)	38.8 (1,351)	42.7 (1,525)
Missing	*50.9 (2,959)*	*46.3 (2,495)*	*22.4 (779)*	*19.6 (699)*

Relational bullying victim, % *(n)*				
Yes	14.1 (817)	17.9 (1,768)	21.2 (738)	25.8 (919)
No	30.3 (1,758)	32.8 (1,768)	51.9 (1,806)	51.7 (1,845)
Missing	*55.7 (3,233)*	*49.3 (2,653)*	*26.9 (937)*	*22.5 (804)*
Self-esteem score, *M (SD)*	19.12 (3.47)	19.34 (3.36)	19.15 (3.48)	19.33 (3.36)
% missing (n)	*45.4 (2,638)*	*40.9 (2,202)*	*5.7 (197)*	*5.7 (203)*
Feeling different, % *(n)*				
Always	2.0 (116)	3.4 (182)	2.6 (89)	3.6 (130)
Mostly	6.0 (349)	8.9 (478)	7.4 (257)	10.8 (387)
Sometimes	21.3 (1,237)	27.6 (1,487)	28.6 (1,235)	33.6 (1,198)
Not often	15.0 (869)	18.3 (986)	19.7 (686)	23.0 (822)
Never	5.0 (291)	4.3 (232)	6.3 (218)	4.8 (170)
Missing	*50.7 (2,946)*	*37.5 (2,019)*	*35.5 (1,235)*	*24.1 (861)*
Feel accepted, % (*n)*				
Agree	35.7 (2,071)	43.6 (2,350)	46.5 (1,619)	53.1 (1,895)
Mostly agree	18.4 (1,067)	20.7 (1,112)	23.7 (826)	24.6 (878)
Mostly disagree	2.8 (161)	2.5 (133)	3.0 (103)	3.0 (106)
Disagree	1.3 (74)	1.4 (76)	1.5 (51)	1.5 (55)
Missing	*41.9 (2,435)*	*31.8 (1,713)*	*25.3 (882)*	*17.8 (634)*
Auxiliary Variables				
Mother's social class based on occupation (32 weeks gestation), % (*n)*				
I (Professional)	3.1 (181)	2.9 (157)	3.6 (126)	3.4 (120)
II (Managerial and technical)	24.5 (1,423)	25.0 (1,346)	26.8 (933)	27.0 (963)
III-NM (Skilled non-manual)	36.1 (2,096)	35.8 (1,926)	35.7 (1,242)	34.7 (1,239)
III-M (Skilled manual)	3.5 (201)	3.0 (162)	3.1 (107)	2.8 (101)
IV (Partly skilled)	14.2 (823)	13.1 (707)	12.8 (444)	11.8 (421)
V (Unskilled)	2.9 (171)	3.2 (172)	2.0 (68)	2.3 (82)

Missing	*15.7 (913)*	*17.0 (914)*	*16.1 (561)*	*18.0 (642)*
Frequency bullied by siblings (12 years 1 month), % (*n)*				
Several times a week	5.6 (325)	7.1 (382)	6.9 (241)	8.5 (302)
About once a week	4.4 (255)	6.6 (356)	5.7 (198)	8.0 (284)
2 or 3 times a month	4.7 (271)	5.7 (308)	6.2 (216)	7.0 (250)
Only ever once or twice	8.5 (491)	11.2 (603)	11.0 (382)	12.6 (449)
Never	28.4 (1,651)	31.1 (1,674)	37.2 (1,295)	27.4 (1,334)
Missing	*48.5 (2,815)*	*38.3 (2,061)*	*33.0 (1,149)*	*26.6 (949)*

Appendix 2

Table A2.1 Results of complete-case linear mixed analyses for boys with masculine-typed behavior as a predictor.

		Emotional Problems				**Peer Problems**			
		Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated
		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)	
Model 1	*β*	-.130	-.076	-.086	-.104	-.174	-.138	-.121	-.136
	95% CI	-.172, -.088	-.136, -.016	-.129, -.043	-.167, -.041	-.216, -.131	-.203, -.072	-.164, -.078	-.203, -.069
	p	<.001*	.014*	<.001*	.001*	<.001*	<.001*	<.001*	<.001*

[a] Higher scores on the PSAI/CAI indicate more masculine parent/self-rated behaviour.

* Significant at the .05 level

Table A2.2 Results of complete-case linear mixed analyses for girls with masculine-typed behavior as a predictor.

		Emotional Problems				**Peer Problems**			
		Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated
		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)	
Model 1	*β*	-.046	.030	-.067	-.125	.083	.177	-.080	-.029
	95% CI	-.093, .001	-.031, .092	-.114, -.020	-.188, -.063	.046, .120	.125, .230	-.116, -.044	-.082, .024
	p	.054	.337	.005*	<.001*	<.001*	<.001*	<.001*	.279

[a] Higher scores on the PSAI/CAI indicate more masculine parent/self-rated behaviour.

* Significant at the .05 level

Appendix 3

Table A3.1 Results of post-imputation linear mixed analyses with for boys with masculine-typed behavior as a predictor. Results presented are beta coefficients (p-value).

		Emotional Problems SDQ				**Peer Problems SDQ**			
		Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated
		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)	
Model 1	Gender-typed behavior	-.122 (<.001*)	-.071 (.004*)	-.083 (<.001*)	-.101 (.002*)	-.155 (<.001*)	-.114 (<.001*)	-.122 (<.001*)	-.124 (<.001*)
Model 2	Gender-typed behavior	-.129 (<.001*)	-.089 (<.001*)	-.075 (<.001*)	-.085 (.007*)	-.165 (<.001*)	-.132 (<.001*)	-.119 (<.001*)	-.119 (<.001*)
Model 3	Gender-typed behavior	-.118 (<.001*)	-.076 (.003*)	-.052 (.011*)	-.070 (.029*)	[b]	[b]	-.103 (<.001*)	-.102 (.001*)
	Self-esteem	-	-	-.008 (.202)	-.017 (.076)	-	-	-.035 (<.001*)	-.041 (<.001*)
	Abuse victimisation	-	-	-	-	-	-	-	-
	Overt bullying victimisation	-	-	-	-	-	-	-	-
	Relational bullying victimisation	-	-	.201 (<.001*)	-	-	-	.374 (<.001*)	.346 (<.001*)
	Feeling different (reference category 'Always')								
	-Mostly	-.167 (.161)	-.218 (.254)	-.073 (.596)	-.156 (.524)				
	-Sometimes	-.3670 (.001*)	-.472 (.007*)	-.347 (.007*)	-.481 (.032*)				
	-Not often	-.486 (<.001*)	-.675 (<.001*)	-.465 (<.001*)	-.730 (.001*)				

-Never	-.585 (<.001*)	-.761 (<.001*)	-.569 (<.001*)	-.849 (.001*)
Feeling accepted (reference category 'Agree')				
-Mostly agree	.192 (<.001*)	.186 (.003*)	.133 (.008*)	-
-Mostly disagree	.443 (<.001*)	.320 (.041*)	.355 (.005*)	-
-Disagree	.682 (<.001*)	.614 (.012*)	.715 (<.001*)	-

Higher scores on the PSAI/CAI indicate more masculine behavior.
Model 1: Unadjusted.
Model 2: Adjusted for demographic confounders.
Model 3: Adjusted for demographic confounders and any covariates which reduced *β* in Step 2 by 5% or more.
[a] No further analysis performed as gender-typed behavior was not a significant predictor of the outcome.
[b] No further analyses performed as no covariates reduced the beta coefficient of gender-typed behavior by 5% or more.
* Significant at the .05 level.

Table A3.2 Results of post-imputation linear mixed analyses with for girls with masculine-typed behavior as a predictor. Results presented are beta coefficients (p-value).

		Emotional Problems SDQ				**Peer Problems SDQ**			
		Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated	Parent-rated	Teacher-rated
		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)		Parent-rated gender-typed behavior (PSAI)		Child-rated gender-typed behavior (CAI)	
Model 1	Gender-typed behavior	-.061 (.005*)	.004 (.902)	-.068 (.005*)	-.121 (<.001*)	.074 (<.001*)	.144 (<.001*)	-.073 (<.001*)	-.032 (.202)
Model 2	Gender-typed behavior	-.064 (.002*)	[a]	-.054 (.020*)	-.099 (<.001*)	.085 (<.001*)	.147 (<.001*)	-.058 (.001*)	[a]
Model 3	Gender-typed behavior	[b]		-.049 (.035*)	[b]	.062 (<.001*)	.127 (<.001*)	-.053 (.002*)	
	Self-esteem			-	-	-.028 (<.001*)	-	-	
	Abuse victimisation			-	-	.232 (<.001*)	.283 (<.001*)	-	
	Overt bullying victimisation			-	-	.225 (<.001*)	.316 (<.001*)	-	
	Relational bullying victimisation			.249 (<.001*)	-	.151 (<.001*)	-	.236 (<.001*)	
	Feeling different (reference category 'Always')								
	-Mostly			-	-				
	-Sometimes			-	-				
	-Not often			-	-				
	-Never			-	-				
	Feeling accepted (reference category 'Agree')								
	-Mostly agree			-	-				
	-Mostly disagree			-	-				
	-Disagree			-	-				

Higher scores on the PSAI/CAI indicate more masculine behavior.
Model 1: Unadjusted.

Model 2: Adjusted for demographic confounders.
Model 3: Adjusted for demographic confounders and any covariates which reduced *β* in Step 2 by 5% or more.
[a] No further analysis performed as gender-typed behavior was not a significant predictor of the outcome.
[b] No further analyses performed as no covariates reduced the beta coefficient of gender-typed behavior by 5% or more.
* Significant at the .05 level.